STIC-ILL

From:	Lacourciere, Karen
Sent:	Monday, December 17, 2001 3:29 PM
To:	STIC-ILL
Subject:	stic ill order

Please provide the following references for use in examining 09/581,331:

Curr. Sci. (2001), 80(2), 161-169

5346351

Can. J. Plant Pathol. (2001), 23(3), 216-235

J. Plant Physiol. (2001), 158(4), 471-478

BioEssays (1989), 10(6), 179-86

Transgenic Res. (2000), 9(4-5), 245-260   1113 1004

Thank-you!

Karen A. Lacourciere Ph.D.
CM1 11D09 GAU 1635
(703) 308-7523

1

A transgenic perspective on plant functional genomics

Andy Pereira
Plant Research International, P.O. Box 16, 6700 AA Wageningen, The Netherlands
(E-mail: A.Pereira@plant.wag-ur.nl)

Abstract

Transgenic crops are very much in the news due to the increasing public debate on their acceptance. In the scientific community though, transgenic plants are proving to be powerful tools to study various aspects of plant sciences. The emerging scientific revolution sparked by genomics based technologies is producing enormous amounts of DNA sequence information that, together with plant transformation methodology, is opening up new experimental opportunities for functional genomics analysis. An overview is provided here on the use of transgenic technology for the functional analysis of plant genes in model plants and a link made to their utilization in transgenic crops. In transgenic plants, insertional mutagenesis using heterologous maize transposons or *Agrobacterium* mediated T-DNA insertions, have been valuable tools for the identification and isolation of genes that display a mutant phenotype. To discover functions of genes that do not display phenotypes when mutated, insertion sequences have been engineered to monitor or change the expression pattern of adjacent genes. These gene detector insertions can detect adjacent promoters, enhancers or gene exons and precisely reflect the expression pattern of the tagged gene. Activation tag insertions can mis-express the adjacent gene and confer dominant phenotypes that help bridge the phenotype gap. Employment of various forms of gene silencing technology broadens the scope of recovering knockout phenotypes for genes with redundant function. All these transgenic strategies describing gene-phenotype relationships can be addressed by high throughput reverse genetics methods that will help provide functions to the genes discovered by genome sequencing. The gene functions discovered by insertional mutagenesis and silencing strategies along with expression pattern analysis will provide an integrated functional genomics perspective and offer unique applications in transgenic crops.

Genomics

Genome sequencing of a number of organisms is providing the scientific community with a vast resource of DNA sequence information that is revolutionizing the way science is being done. Due to technological breakthroughs DNA sequence information of small and medium sized genomes is being uncovered 'en masse'. An organism that used to be studied at the individual gene level can now be examined in terms of its genome organization, expression and interaction – a field that has been given the new buzz word name 'genomics'. The science of genomics is the study of how genes and genetic information are organized within the genome, and how formation are organized within the genome, and how

this organization determines their function. This science was given an impetus by a bold plan to sequence the human genome that stimulated the development of efficient and cheap high throughput (HTP) sequencing techniques. A large number of microbial genomes are being sequenced, with six archaea and 17 bacteria already in public databases while the genomes of six more bacteria are completed. Among the eukaryota the complete genomes of yeast and *Caenorhabiditis* have been published (Mewes et al., 1997; The *C. elegans* sequencing consortium, 1998) while that of *Drosophila* is recently reported to be complete. Among higher eukaryotes six complete chromosomes have been published including one human and two *Arabidopsis* chromosomes. The complete genomes have

already entered the post-genomic era where gene functions are being systematically uncovered by functional genomics methodology.

Model genomes

Plant genomics has taken root with the establishment of model plants like *Arabidopsis* that set a large group of scientists to dissect this plant using the tools of genetics, biochemistry, physiology and molecular biology. *Arabidopsis* is a weed that grows all over the world, has one of the smallest plant genomes (140 Mb, or million base pairs), is easy to grow, transform and use for genetic experiments. Structural genomics research starting from genetic and physical maps to HTP whole genome sequencing, is a publicly funded multinational effort from research groups in Japan, Europe and the United States. The sequences of chromosome 2 (19.6 Mb) and chromosome 4 (17.38 Mb) have recently been published (Lin et al., 1999; Mayer et al., 1999) and the whole genome sequence will be available by the end of this year. The *Arabidopsis* genome sequence provides us with a blueprint of the basic set of genes that is required for the growth and maintenance of a simple plant, but the systematic correlation of gene sequence and specific function in the plant still needs to be done. Some of the remarkable insights emerging from whole chromosome sequences are the number of gene clusters, large-scale duplications of chromosome segments, insertions of huge chunks of mitochondrial DNA, as well as 40% of newly discovered genes of unknown function. While the international *Arabidopsis* sequencing effort is concentrating on sequencing the ecotype Columbia, genome sampling or random sequencing of another ecotype, Landsberg erecta, by groups including TIGR that has 15,000 sequences on its website (www.tigr.org), will contribute many markers including SNPs (single nucleotide polymorphisms) for fine mapping of mutants. The similarity of genes and gene order between *Arabidopsis* and its closest relatives the crucifers (Gale & Devos, 1998) will facilitate the direct use of *Arabidopsis* genomic and functional information in the crop plants.

While *Arabidopsis* is very much the model dicot plant, rice with its relatively small genome (450 Mb) has been selected to be the model monocot plant whose genome will be sequenced by a multinational effort. The other cereals like maize, barley and wheat have genomes of size 3000, 5000 and 16000 Mb, respectively that will not soon be sequenced completely.

The demonstration that genetic selection for gene rich sequence clones can be made for maize genome sequencing (Rabinowicz et al., 1999), suggests that shotgun sequencing of the gene rich part of large genomes will be a feasible method for gene discovery. The gene order or synteny between genes in the monocots is well conserved (Gale & Devos, 1998) so that the compact rice genome sequence will help to decipher the gene order and content in the more complex genomes. As the monocot model is also a crop plant, with a half billion tons grown annually worldwide and 90% used for consumption, deciphering this model becomes attractive for commercial applications. An additional genome model gaining recognition is the legume *Medicago truncatula* to dissect symbiotic nitrogen fixation (Nam et al., 1999). Besides the model genomes, some plants may be considered models due to their characteristic advantages, for example, maize genetics, tobacco cell culture, tomato fruit biology. The accessibility of genomic biology approaches to these and genome models will help our understanding of plant biology from a problem oriented (bottom-up) approach as well as the model genome (top-down) approach.

EST sequencing

The expressed genes of a plant can be effectively discovered by sequencing expressed sequence tags (ESTs or complementary DNA from transcribed genes) that reveal the active genes in the genome in a specific tissue. Plant EST collections of about 160,000 are available in public databases from 19 plant species like *Arabidopsis*, rice, tomato, maize, soybean, cotton, loblolly pine etc, and offer an efficient way of gene discovery in these plants. The ESTs are redundant with highly expressed genes represented many times so sequence comparisons are required to quantify the number of genes.

There are 37,746 *Arabidopsis* ESTs (Höfte et al., 1993; Newman et al., 1994) submitted to dbEST, and after comparisons to characterized expressed and predicted transcripts, a set of about 16,000 unique transcripts has been identified in the *Arabidopsis* Gene Index (Quackenbush et al., 2000). Previous analysis (Rounsley et al., 1996) suggested that ESTs exist for around 55–60% of the *Arabidopsis* genes, and this is confirmed with the present genomic sequencing results. EST comparisons could also be used to define unique tags to individual members of multigene families like cytoplasmic ribosomal proteins (Cooke et al.,

Material may be protected by copyright law (Title 17, U.S. Code)

ich
se-
hat
en-
he
ots
the
her
en-
int,
ind
del
An
the
ni-
del
lue
ize
gy.
to
ing
up)
vn)

lis-
iTs
hat
ific
are
:ies
on,
ene
ant
nes
the

al.,
ST,
ind
jue
ene
'sis
for
i is
'es-
ine
un-
al.,

1997) where 50 different proteins are encoded by 106 genes.

For other plants dbEST contains 46,268 rice, 29,209 maize, 26,255 tomato and 17,738 soybean ESTs that have been classified into individual Gene Indices (Quackenbush et al., 2000). Surprisingly though, proprietary databases, for example from DuPont, contain 127,247 wheat, 389,733 maize, 114,162 rice and 183,663 soybean ESTs (Meyers et al., 1999). These efforts lend support to the idea that for gene discovery in the larger genomes we need not have to wait for genome sequencing. To avoid the redundancy of random EST sequencing, normalized cDNA libraries have been used to create a 'unigene' set of clones representative of the genome.

A comparison of EST databases from different plants and tissue/conditions reveals the diversity in coding sequences between plants, yet provides a global perspective of the similarities in genes for specific tissues (e.g. ripening) or conditions (e.g. pathogen induced). The sequence similarity between the plant genes, analyzed by bioinformatics tools, permit assignment of probable gene function and identification of orthologs between species. Even though similarity screens to databases can assign some degree of general function to genes for transcription factors or signal transduction, their exact function requires experimental approaches.

Functional genomics approaches

The field of functional genomics has blossomed forth to specifically address the function of genes discovered by genome sequencing. In contrast to the previously prevalent gene by gene approaches, new HTP methods are being developed for expression analysis as well as the recovery and identification of mutants. The experimental approach is consequently changing from a hypothesis driven procedure to a non-biased data collection and archiving methodology that makes these data available in relational databases that can be analyzed later by bioinformatics tools. The functional genomics methodology is also changing the experimental strategy from a Forward Genetics, that is, mutant to gene approach, to a Reverse Genetics or sequenced-gene to mutant and function approach (Bouchez & Höfte, 1998). Though still at its infancy, functional genomics of model plants is expected to bear fruit towards the understanding of basic plant biology, as well as the exploitation of genomic information for crop plant improvement. This is because a

large number of gene-functions for generic traits will be functional across species, either directly or after identifying the functional homologues.

Expression analysis

The major developments in whole genome analysis have been in the field of transcript expression analysis using a variety of HTP methods. In plants different technologies have been employed based on high-density nylon filters (Desprez et al., 1998), microarrays (Schena et al., 1995), various gel systems (Baldwin et al., 1999) and even a sequencing-based method SAGE (Matsumura et al., 1999). Some novel systems that offer much promise are a new gel based strategy called GeneCalling by CuraGen Corporation (Bruce et al., 2000) and a revolutionary microbead system from Lynx Therapeutics (Brenner et al., 2000). The most popular are expression microarrays on chips due to the availability and convenience of the technology that is also applicable to non-model plants (Lemieux et al., 1998). With gene chips or microarrays all the genes of an organism (as ESTs or synthesized oligonucleotides) can be placed on a solid support like glass and used in hybridization experiments with different RNA samples to reveal the gene expression patterns. This is quite commonly used in yeast but progressively being employed in more complex organisms, ranging from *Arabidopsis* (Ruan et al., 1998) to maize and humans. The expression profile of a developmental stage or induced condition can help identify genes and coordinately regulated pathways and their functions deduced by association. However, the expression patterns of genes supply correlative information and do not necessarily prove a causal relationship between gene sequence and function. As factors other than mRNA level alone, for example, protein modification or metabolite flux, determine the activity of a gene product *in situ*, even genome wide expression analysis cannot prove a gene function relationship and other supplementary experimental data is required.

The real picture of gene expression is offered by proteomics (Dove, 1999) that addresses the protein expression of a cell type, the protein modifications and can also reveal actual interactions in the cell. Recent developments in mass spectrometry have simplified protein analysis and characterization and have become automated. The availability of whole genome sequences in databases provides the basis for the identification of peptide tags from analyzed pro-

teins. The proteome or protein content of a cell type, is limited to normally expressed proteins while minimally expressed products are beyond the limits of resolution, probably requiring separate enrichment or sensitization methods. As the proteome is also quite complicated to analyze for plant cells with their cell walls, the metabolome provides a picture of all metabolites that are present in the cell/tissue type revealing the active biochemical processes. Methods for plant metabolite profiling (Trethewey et al., 1999) are still at infancy and due to the nature of the system are very prone to uncontrolled variation. Further technological developments along with genome analysis information hold great promise for the analysis of the plant proteome and metabolome. Expression analysis tools combined with mutant or over-expression analysis of genes will then be able to provide a unique multi-dimensional picture of genetic circuits and pathways.

Gene-phenotype relationship

The phenotype associated with a gene function is often the best clue to its role in the plant. Mutant phenotypes can be broadly defined at the morphological, biochemical or physiological levels and provide information on the interaction between different processes. Classically, chemical and physical mutagens have provided loss of function recessive mutants and have helped define genes involved in specific pathways or processes. To study the interaction between genes, mutants in a particular pathway can be combined and the genetic hierarchy studied. The genetic circuits of simple processes like pigmentation have been defined in a number of plants (Mol et al., 1998) while others like the ethylene response pathway (Solano & Ecker, 1998) are best studied in models like *Arabidopsis*.

The genetic analysis of mutants requires good plant systems that are easy to handle and are genetically simple with minimal gene redundancy. *Arabidopsis*, maize, tomato, petunia and barley have served as good systems to isolate mutants using chemical or physical agents and map the corresponding loci on chromosomes. These mutants can then be precisely mapped and lead to gene isolation by map-based cloning procedures from large YAC/BAC libraries. Contigs or physical maps of YAC/BAC clones are available for *Arabidopsis* (Schmidt et al., 1995) and this facilitates the identification of the corresponding gene. Once the gene has been identified, complementa-

tion of the mutant by transformation (Arondel et al., 1992) is the most straightforward way of establishing the gene-phenotype causal relationship. The proof of isolation of the gene corresponding to the mutant can also be demonstrated by sequencing a number of alleles (Shirasu et al., 1999) in cases where complementation of the mutant by transformation is not easy.

Two features determine the ease of gene identification by map based cloning approaches: the availability of linked markers and transformation protocols. Markers can be generated at high efficiency in any plant by methods like AFLP using bulked segregation analysis. In *Arabidopsis* the generation of SNPs and HTP screens in segregating populations (Cho et al., 1999) can aid in determining the precise map position within a few weeks. Together with efficient transformation systems like the floral dip method in *Arabidopsis* (Clough & Bent, 1998) gene-function relationships with mutants can be established in a semi HTP manner. Consequently, with the help of the genome sequence, the genes corresponding to all the known mutants can be identified in a cost-effective way.

Functional genomics with heterologous insertion sequences

Knockout mutagenesis

Insertional mutagenesis for gene identification offers several advantages for functional genomics over mutants derived from chemical/physical treatments. Two types of insertion sequences are commonly used in plants: transposable elements and *Agrobacterium tumefaciens* mediated T-DNA (transfer DNA) insertions. The insertion sequence often causes a knockout mutation (Figure 1B) by blocking expression of the gene and might display a mutant phenotype. The mutant gene that is tagged by the insertion sequence can then be isolated by recovering DNA flanking the insert and subsequently lead to the isolation of the wild-type gene sequence. This method of gene tagging with well-characterized insertion sequences as molecular tags does not need prior knowledge of the gene product or expression. The advent of efficient insertional mutagenesis strategies, first with the endogenous transposons in maize, later with T-DNA in the model *Arabidopsis* and heterologous transposons in a number of good genetic systems has rapidly facilitated gene identification and isolation (Pereira, 1998).
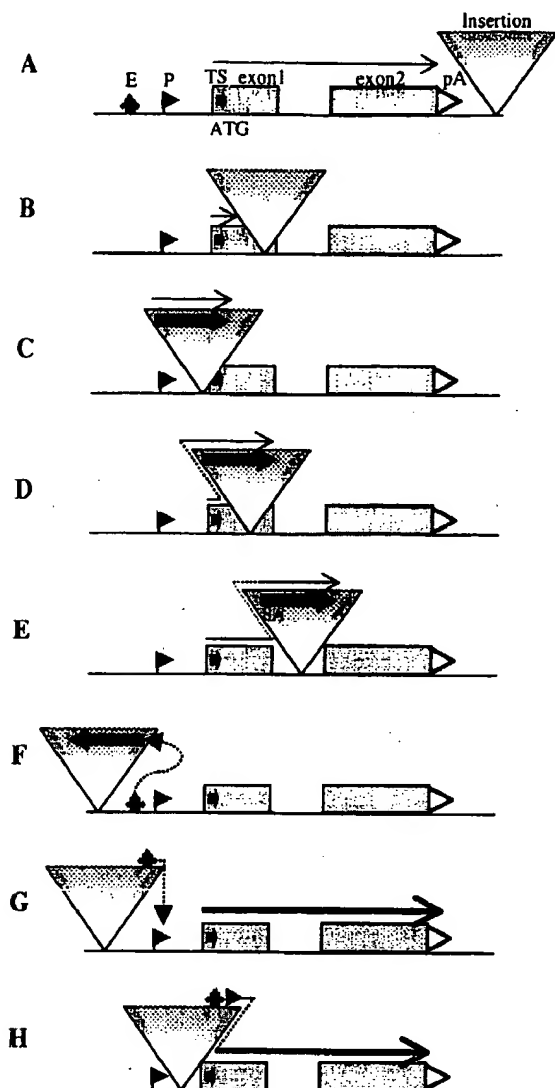
al.,
.sh-
oof
:ant
.ber
>m-
not

.ca-
lity
.irk-
.ant
.na-
and
al.,
po-
.ent
l in
re-
:mi
en-
the
:ive

D

of-
.ver
nts.
sed
:um
;er-
.out
the
[he
nce
the
the
ag-
as
the
.ent
do-
the
n a
.ted

Figure 1. Insertion sequences for gene function identification. Loss-of-function mutations are possible with insertion types B–E, Gain-of-function mutations with types G & H, Gene expression detection with types C–F. (A) Wild-type gene structure showing different components with symbols: E-enhancer, P-promoter, TS-transcription start, thick arrow- translation start ATG, exon 1 & 2 of standard gene as shaded box with intervening intron, pA-transcription termination site, insertion- T-DNA or transposon as triangle insert in gene. Thin arrow indicates gene transcription; thick arrow indicates reporter gene within insertion sequence. (B) Knock-out insertion within coding region of gene blocking transcription. (C) Promoter trap with insertion-reporter under transcriptional control of tagged gene. (D) Gene trap with insertion reporter lacking translation start, as translation fusion with tagged gene. (E) Exon trap (gene trap) with insertion reporter bearing a splice acceptor SA to make gene fusions with tagged gene. (F) Enhancer detection with Insertion Reporter displaying flanking gene enhancer activity. (G) Activation tag with insertion carrying a strong enhancer that activates the expression of adjacent genes. (H) Promoter read-out insertion with a strong promoter reading out into adjacent gene causing mis-expression.

The identification of specific mutants is the primary rate-limiting step but consequently leads to the identification of the gene. The most disturbing feature of an insertional mutagenesis strategy is that the process of transformation often causes 'somaclonal' mutations that are not tagged by the T-DNA insertion. This frequency of non-tagged mutants is quite high in tissue culture transformation systems as was observed for tomato (Yoder, 1990) and *Arabidopsis* (Koncz et al., 1992). The major advance in the use of T-DNA tagging was made in *Arabidopsis*, where non-tissue culture transformation approaches like seed transformation (Feldmann, 1991), vacuum infiltration (Bechtold & Pelletier, 1998) or floral-dip (Clough & Bent, 1998) methods displayed a much lower proportion of untagged mutants. In any case, these forward genetics strategies for recovering T-DNA insertion mutants require careful genetic analysis and proof by complementation after transformation. For forward genetics, T-DNA insertional mutagenesis is only practical on a large scale for *Arabidopsis* and has resulted in a large number of tagged genes. Molecular analyses of tagged mutants (Azpiroz-Leehan & Feldmann, 1997) have shown that insertions can occur all over a gene, even insertions in introns and promoters revealing mutant phenotypes. Moreover, the tagged genes are distributed randomly over all chromosomal regions. For reverse genetics strategies, many T-DNA insertion populations have been developed in *Arabidopsis* (See Table 1), but are equally applicable in other plants like rice with efficient transformation procedures. Another knockout mutagenesis tool in rice is an endogenous retrotransposon *Tos17* that has been developed for reverse genetics screens (Hirochika, 1997).

Transposon tagging has been an effective strategy to identify genes in plants like maize with its well-studied endogenous transposons (Walbot, 1992). The maize transposon systems *Ac–Ds* (Baker et al., 1986) and *En-I(Spm)* (Pereira & Saedler, 1989; Masson & Fedoroff, 1989), have been shown to transpose after being introduced into numerous heterologous hosts. These heterologous transposons can be modified *in vitro* and thus offer several advantages for transposon mutagenesis over the endogenous systems (Pereira, 1998). To develop effective tagging strategies, two component systems, comprising a mobile transposon component (*Ds* or *I/dSpm*) and the corresponding stable transposase (*Ac* or *En/Spm*) source, have been subsequently used (Bancroft et al., 1992; Aarts et al., 1993). The mobile transposon components are engin-

Table 1. Insertion sequence tags used for reverse genetics in *Arabidopsis*

Tag	Screenable marker	# inserts per plant	Insert population	PCR selection	Insert sequencing	Reference
T-DNA	NPTII	~1.5	5,300 lines	x		McKinney et al., 1995
T-DNA	NPTII	Low	9,100 lines	x		Krysan et al., 1996
T-DNA	NPTII	~1.5	6,000 lines	x		Winkler et al., 1998
T-DNA	PT- NPTII	Low	Cell culture		x	Mathur et al., 1998
T-DNA	ET-GUS	Low	11,370 lines			Campisi et al., 1999
T-DNA	NPTII	Low	60,480 lines	x		Krysan et al., 1999
T-DNA		1-low	9,264 lines	x		Meissner et al., 1999
DsG/DsE	NPTII	1-low	2,000 lines		x	Martienssen, 1998
En-1	–	~6	8,000 lines	x		Wisman et al., 1998
I/dSpm	–	~20	2,592 lines	x	x	Speulman et al., 1999
dSpm	BAR	1-low	48,000 lines	x	x	Tissier et al., 1999
DsG	NPTII	1-low	931 lines		x	Parinov et al., 1999

PT – Promoter trap; ET – Enhancer trap; GT – Gene trap.

eered to carry various marker or selectable genes like antibiotic or herbicide resistance genes. In order to monitor or select their excision transposons are often inserted in other assayable/selectable marker genes (Baker et al., 1987; Jones et al., 1989). In attempts to control transposition, the transposase is often put under the regulation of heterologous promoters (Swinburne et al., 1992), or segregated out in progeny to yield stable transposon inserts. The first few heterologous tagged mutants in *Arabidopsis* (Aarts et al., 1993; Bancroft et al., 1993; Long et al., 1993) and petunia (Chuck et al., 1993) were obtained by a random tagging strategy, in which plants containing transposed elements were selfed and screened for obvious mutant phenotypes. The frequency of random inserts displaying a mutant phenotype is about 1–5%, depending on the screening strategy used, similar to that observed for T-DNA inserts.

As transposons move preferentially to closely linked sites (Jones et al., 1990; Bancroft & Dean, 1993), this feature is useful to efficiently isolate mutants for genes located near the original position. In heterologous systems this is achieved by mapping a large number of transformed T-DNA inserts that serve as jumping pads for local mutagenesis. This strategy was effectively demonstrated in tomato where 'targeted tagging' with a *Ds* transposon linked about 3 cM away from the target gene yielded tagged mutants at a high rate of about 1/1,000 transpositions (Jones et al., 1994). Similar directed tagging strategies have since been used to tag genes in some of the larger plant gen-

omes and have made this strategy a practical reality for gene isolation.

Gene detector insertions

Most established genetic descriptions of biological processes come from loss-of-function or knockout mutations. Yet, in most higher organisms the majority of genes display no obvious knockout phenotype (Burns et al., 1994). Part of the reason for this 'phenotype gap' could be functional redundancy, where one or more other loci (mostly homologous) can substitute for the same function. In fact even in *Arabidopsis* with its simple genome large chromosomal segments are duplicated (Lin et al., 1999; Mayer et al., 1999) comprising redundant genes. Sequential disruption of homologous and redundant genes in an individual genotype might ultimately reveal a mutant phenotype. Some subtle mutants might require a thorough screening system for their detection, while others have conditional phenotypes and are revealed only when challenged with appropriate environmental cues like pathogens or abiotic stress. Quite often a mutant phenotype reflects only a part of the gene function, primarily the first expressed defect, for example, in embryo lethal mutants that may have wide-ranging functions from development to metabolism. Even in the simple model *Arabidopsis* the frequency of recovered lethal mutants is low when compared to other organisms (Miklos & Rubin, 1996), suggesting that a large proportion of gametophytic knockout mutants

might be difficult to recover. All these reasons suggest that other gene detection systems rather than standard knockouts are required for the identification of gene function.

The principle that expression patterns are useful parameters to assign gene function is extensively used in the application of expression microarrays and related technology. However, microarray based technology does not have the resolution to monitor the expression of specific genes within single or few cells. Entrapment strategies (Skarnes, 1990), outlined in Figure 1, make use of inserts containing reporter gene constructs, whose expression is dependent on transcriptional regulatory sequences of the adjacent host gene. In this way genes (adjacent or knockouts) can be identified by their expression pattern, even though they might not directly display an obvious mutant phenotype. The identification of very specific expression patterns, for example, in the meristem or localized after pathogen induction, can suggest the possible phenotype to screen for. The entrapment inserts allow for the selection of inserts in specific classes of genes, based on their expression pattern. Subsequent production of double or multiple mutants in a pathway or between partially redundant genes, as indicated by the expression pattern, might finally reveal mutant phenotypes.

Enhancer detection was developed to detect enhancers in the genome, which are capable of orientation independent transcriptional activation from a distance, and thus identify genes based on their expression pattern. Enhancer detection constructs (Figure 1F) contain a reporter gene like β-glucuronidase (GUS) with a weak or minimal promoter, for example with a TATA box situated near the border of the insert. When integrated in the vicinity of an enhancer sequence in the genome, the reporter gene can display the expression pattern of the nearby chromosomal gene. The inserts will be activated from a distance and independent of the orientation of the reporter to the tagged plant gene, so the frequency of enhancer detector insertions displaying some expression pattern is as high as 50%. One disadvantage of these inserts is that it is difficult to pinpoint the exact location of the tagged gene that might be located at an unpredictable distance especially in large complex genomes. In a compact or completely sequenced genome like *Arabidopsis* the gene prediction and identification should be easier.

'Gene trap' type inserts (Figure 1D) are designed to create fusion transcripts with the target gene

(Skarnes, 1990). In plants the reporter genes *nptii* and *gusA* are able to support gene fusions and have been used widely in T-DNA vectors as gene fusion traps (Topping & Lindsey, 1995). Promoter trap types (Figure 1C) consist of a promoterless reporter gene and are expressed when inserted downstream of the chromosomal gene promoter. The exon trap type (Figure 1E) is more versatile as it enables reporter gene fusions to be created at various locations within a gene. By introduction of splice acceptor sites upstream of the reporter gene, transcriptional fusions are created even for insertions in introns and thus increases the frequency of inserts expressing the reporter gene. Reporter gene fusions can help in localization of the gene product, for example, in the nucleus, organelle or membrane.

The first T-DNA vectors were designed as promoter traps (Koncz et al., 1989) with the *nptii* reporter gene. Later the *gusA* reporter gene was employed for enhancer and promoter traps (Kertbundit et al., 1991; Topping & Lindsey, 1995). These studies revealed that a high proportion of inserts displayed reporter gene expression, with about 25% for promoter traps and 50% for the enhancer traps. Most surprisingly the frequency of expressed inserts was similar in *Arabidopsis* and larger genomes like tobacco (Koncz et al., 1989), suggesting that T-DNA inserted preferentially in transcriptionally active regions. Translational gene fusion vectors (Figure 1D) with the *nptii* gene lacking a translation start and placed near the T-DNA border (Koncz et al., 1989; Babiychuk et al., 1997) yielded kanamycin resistant calli/seedlings and also resulted in the identification of gene knockout mutants. There are large T-DNA collections being produced now in *Arabidopsis* that employ promoter trapping (Table 1). These include the vacuum infiltration derived Versailles T-DNA collection (Bechtold & Pelletier, 1998), a collection of 11,370 (Campisi et al., 1999) lines available through stock centers, and a cell culture transformed population (Mathur et al., 1998) that has been analyzed by sequencing. Although large collections of T-DNA transformants for enhancer and promoter traps have been produced, analyzed, and scores of interesting patterns identified, it has not been easy to make the exact correlation between the reporter expression pattern and the resident gene. One reason could be due to the presence of spurious enhancers and promoters in the genome that are not associated with the expression of the adjacent genes but are picked up by these expression detectors. Another reason is the difficulty in determining the position of the ex-

252

pressed trapped gene, especially in enhancer traps and with rearranged or multiple T-DNA inserts in complex genomes.

For transposons similar strategies were employed for the *Ac–Ds* transposons engineered as promoter and enhancer trap systems. A *Ds*, carrying a CaMV minimal promoter upstream of the *gusA* reporter gene and a hygromycin resistance marker, was mobilized by an *Ac* transposase with linked *iaah* gene as a negative selection marker (Fedoroff & Smith, 1993). With these selection markers stable transposed *Ds* elements could be recovered in the progeny. A novel gene *LRP1* expressed in root development (Smith & Fedoroff, 1995) was identified that displayed no mutant phenotype, probably due to gene redundancy. An embryo-defective lethal mutation was also identified (Tsugeki et al., 1996) whose expression could be analyzed in heterozygotes, demonstrating the use of enhancer entrapment for function identification.

A novel method for efficient selection of stable transpositions has been developed for an *Ac–Ds* based gene (exon) trap and enhancer trap system (Sundaresan et al., 1995). From the progeny of starter plants carrying heterozygous *Ac–Ds* elements, stable transposed *Ds* elements carrying the *nptii* gene are selected using the negative selectable *iaaH* marker gene for segregating out the transposase. Independently derived *Ds* inserts were screened for their GUS expression pattern and revealed some GUS reporter gene activity in about 50% enhancer trap (*DsE*) and 25% gene trap (*DsG*) inserts (Sundaresan et al., 1995). The first gene identified by gene trapping, *PRO-LIFERA* a gene expressed in megagametophyte and embryo development (Springer et al., 1995), exemplifies the utility of the system. A number of starter *DsG* and *DsE* lines have since been used by different collaborators to produce about 10,000 stable transposed elements, many of which have been screened for their expression patterns (Martienssen, 1998). A similar strategy for selection of stable transposed gene trap *DsG* elements has been initiated in rice (Chin et al., 1999) using greenhouse selectable marker systems.

Mis-expression mutants

The classical mutational spectrum using chemical/physical mutagens or knockout insertions reveals a low frequency of mutant phenotypes probably due to redundant gene functions, conditional and subtle phenotypes. Gain-of-function mutants by the spatial or temporal mis-expression of individual genes provides an alternative way to perturb gene function or regulatory networks. One way this can be achieved is by employing insertion sequences that carry a strong enhancer element near the border, thus activating the adjacent gene ectopically or just enhancing its expression. The recovery of over/mis-expression mutants may be advantageous in cases where positive selection is possible for dominant mutations, or in processes or genes where simple knockout mutants reveal no mutant phenotype.

This method of 'Activation tagging' has been applied successfully in *Arabidopsis* (Kakimoto, 1996), with a T-DNA vector containing multiple strong CaMV enhancer sequences near the border (Figure 1G), and selecting transformants for cytokinin independent regeneration. A similar activation tagging screen was used for the identification of an early flowering mutant that turned out to be due to activation of a gene *FT* (Kardailsky et al., 1999), which was characterized independently (Kobayashi et al., 1999) as a knockout mutant with a late flowering phenotype. The dissection of genetic pathways using second-step mutants, was demonstrated by the selection of a suppressor of a phytochrome *phyB* mutant with activation tagging (Neff et al., 1999). The identified phenotype was shown to be caused by enhanced expression of a gene *BAS1* involved in Brassinosteroid levels, and demonstrates a link between the light response and hormone signaling pathways.

An extensive description of screening of more than 25,000 T-DNA activation tags has been published recently (Weigel et al., 2000), revealing a 1/1000 frequency of confirmed dominant mutants. This frequency is surprisingly much lower than that observed for recessive knockout mutant screens, suggesting that there is either a bias in the activation tag insertional spectrum or limitations (e.g. insulators) to the interaction of the introduced enhancer and host promoters. The dominant mutants were shown to be caused by presence of the enhancer tag 0.38–3.6 kb from the overexpressed gene. These gain-of-function mutants provide new variants for gene function analysis.

A transposon construct variant (Figure 1H) to isolate dominant gain-of-function alleles employed the CaMV 35S promoter transcribing outward from the *Ds* transposon end. The first gene tagged with this system was *TINY* (Wilson et al., 1996), which was recovered as a semidominant overexpression mutant. Later another semidominant dwarf mutant was obtained that

exhibits reduced gibberellin response due to overexpression of *SHI*, a regulatory gene containing a zinc finger region. By this system, dominant mutations might be caused by insertions in the transcription unit over/mis-expressing the gene. But it might be expected that insertions producing an anti-sense transcript also produce dominant mutants due to suppression. As the frequency of overexpression inserts is not expected to be higher than knockout mutations by the same transposon, both types of mutations will be obtained in a screen.

Site specific deletions

Specific genomic deletions can help characterize the function of individual genes and other chromosomal structural elements like matrix associated regions or general enhancers in the genome. In the light of the finding of genetic redundancy due to gene family clusters on the sequenced chromosomes (Lin et al., 1999; Mayer et al., 1999), the ability to create small chromosomal deletions encompassing a set of the duplicated genes can contribute to gene function analysis of these multigene families.

Homologous recombination has not been very efficient in plants to create site specific mutations and deletions. Heterologous site specific recombination systems like the bacteriophage P1 *Cre-lox* and the yeast *FLP–FRT* systems have been shown to work in plants to create deletions, and in combination with transposition can be used for gene identification (van Haaren & Ow, 1993). Recombination occurs between two *lox* sites or *FRT* sites, mediated respectively by the *Cre-* or *FLP-* recombinase.

Constructs have been used (Osborne et al., 1995) that contain the *lox* recombination sites, both within the transposon (*Ds-lox*) and the adjacent T-DNA. After transposition of *Ds-lox* to a closely linked site, the sequence specific recombination sites are situated close together on the chromosome, so that recombinase mediated small deletions can be recovered (by *Cre*-recombinase line crosses). Analysis of deletions has suggested that the transposed *Ds* elements were not distributed randomly but in clusters. The potential site specificity of insertions and the non-recovery of large deletions (Stuurman et al., 1998) may thus bias the generation of random deletions all over the genome.

Reverse genetics strategies

Site selected insertional mutagenesis

Reverse genetics strategies in yeast can use high frequencies of gene replacement to inactivate genes or putative ORFs, an option not available for plants. In *Drosophila* and *Caenorhabditis*, a reverse genetics technique called 'site selected insertion mutagenesis' was developed using transposons (Ballinger & Benzer, 1989; Kaiser & Goodwin, 1990, Zwaal et al., 1993) to inactivate genes that had been sequenced, but whose function was unknown. In this strategy insertion mutants of the entire genome are first generated and then individuals with an insertion in a gene of interest identified by PCR techniques. In maize, snapdragon and petunia (Walbot, 1992; Coen et al., 1989; Gerats et al., 1990) the multiple endogenous transposon copies provide genome saturation and site selected insertional mutagenesis has been shown to work to identify knockout mutations in specific genes (Das & Martienssen, 1995; Koes et al., 1995). As an alternative to insertions, deletion mutants generated by chemical or physical mutagens can be employed in a HTP screen to identify mutants of specific genes described by the genome sequence (Liu et al., 1999).

The chance of recovering an insert in a target gene is dependent on the population of inserts, the size of the genome and the size of the gene target. Using transgenic systems large populations of inserts can be generated by T-DNA or transposon inserts. These can be knockout insertions as well as the variety of gene detection and mis-expression inserts described above that will help gene function analysis. Currently two strategies for genome saturation with inserts are being pursued:

(i) single stable elements, similar to the T-DNA population structure, and

(ii) multiple elements from active transposing populations, similar to the endogenous transposon systems.

In *Arabidopsis* about 110,000 inserts would give an insert every kb, with about a 99% chance to mutate an *Arabidopsis* gene of 5 kb (Krysan et al., 1999). Larger genomes being investigated like rice, tomato and maize would require sizeable larger number of inserts unless the frequency of insertions in genes is higher.

The selection of defined gene mutations is done by PCR using a pair of primers, one of which anneals to the insertion sequence and the other to the specific

254

target gene. If an insert close to the primer is present in the population then a PCR product specific for the target gene can be observed. These sensitive PCR screens can be scaled up to HTP so that by specific pooling strategies the individuals containing the insert can be directly identified from a large population. Progeny of the individual containing the insert can then be analyzed to characterize the mutant phenotype.

In *Arabidopsis* at first a T-DNA population of 5300 transformants with about 8000 inserts, was screened in 53 pools of 100 transformants each, to identify insertions in two members of the actin multi-gene family (McKinney et al., 1995). Later, a hierarchical pooling strategy with DNA from 9100 transformed lines was used to recover inserts in 17 out of the 63 signal transduction genes tested (Krysan et al., 1996). These reverse genetics screening techniques have been used to isolate and characterize mutants, providing new insights into gene function and validating the strategy (Hirsch et al., 1998; Gaymard et al., 1998). At present a collection of 60,480 transformants is publicly available for screening by PCR through the *Arabidopsis* Knockout Facility at the University of Wisconsin (Krysan et al., 1999). The pooling strategy and organized screening done on available DNA involves a primary PCR screen on 30 DNA pools and after identification of a positive primary pool, a secondary PCR screen. This identifies a pool that has to be sown out by the experimenter, DNA isolated and the individual containing the insert identified.

There are various transposon populations of the *En-1 (Spm–dSpm)* system available at present in *Arabidopsis*, outlined in Table 1, that are also accessible to PCR screening by similar methods as described for the T-DNA pools (Wisman et al., 1998; Speulman et al., 1999; Tissier et al., 1999). Two of the populations contain multiple actively transposing elements per line that have been accumulated by propagation of the lines for a number of generations (Wisman et al., 1998; Speulman et al., 1999). A large stable *dSpm* insert population has been recovered by an ingenious system (Tissier et al., 1999), using herbicide selectable markers in the greenhouse for positive (*bar* gene conferring Basta resistance) and negative selection using the *SU1* gene (O'Keefe et al., 1994).

In contrast to the often-complex nature of T-DNA insertions, transposon inserts have a defined terminal sequence rendering them all accessible to PCR selection. In addition, due to their ability to jump to adjacent sites, transposon insert derived mutations can also be generated for members of closely linked gene

families. In fact, the local transposition behavior of the *Ac–Ds* transposons has been applied to saturate specific genomic regions with inserts (Dubois et al., 1998; Ito et al., 1999). Thus mapped transposons in the genome are valuable as donor sites to produce saturated populations of inserts in the linked chromosomal regions.

The reverse genetics strategies using PCR screens outlined above can be applied for insertion populations that are large enough to be able to recover inserts in specific genes of interest. Alternatively, the DNA flanking an insert can be isolated, sequenced and compared to the genome sequence. There is a variety of PCR based methods to isolate genomic DNA flanking inserts (reviewed in Maes et al., 1999) that can then be sequenced. After sequencing a large number of these insertion sites, inserts in genes of interest, can be identified and their mutant phenotype analyzed. With more than 50% of the *Arabidopsis* genome sequence available, a number of groups have sequenced DNA flanking transposon inserts and been able to identify tagged genes very efficiently (Ito et al., 1999; Parinov et al., 1999; Speulman et al., 1999; Tissier et al., 1999). These efforts are also very valuable as an unbiased system to test insertional specificity and also for the characterization of inserts in or flanking genes that have not yet been predicted by bioinformatics, but are an accumulating resource towards insertions in all genes. The value of sequencing transposon-flanking DNA has also been recognized in the larger genome rice where active *Ac* transposon populations have been used for the identification of inserts in genes by sequencing in parallel with PCR based screening (Enoki et al., 1999).

Gene silencing

Homology dependent gene silencing (HDGS) using transgenes offers a new tool to shut down the function of endogenous genes, consequently generating mutants without mutation. Two types of HDGS have been described (Kooter et al., 1999); transcriptional gene silencing (TGS) involving promoter homology/methylation and post-transcriptional gene silencing (PTGS) requiring homology between interacting genes in the transcribed regions. At first PTGS was found associated with tandem/inverted repeats or methylated and minimally expressed transgenes. To obtain gene silencing most experiments typically used a complete or partial cDNA clone placed downstream of a strong promoter in sense or anti-sense

255

r of
irate
: al.,
1 the
itur-
)mal

eens
ula-
serts
)NA
om-
y of
ank-
can
er of
n be
With
ence
)NA
itify
inov
al.,
un-
also
enes
, but
1 all
king
ome
een
se-
ioki

us-
the
rat-
)GS
ran-
oter
ene
iter-
'GS
eats
nes.
ally
wn-
nse

orientation. In about 10–50% of the transformants a dominant mutant phenotype was observed for genes with a visual effect, for example, developmental regulatory genes, or enzymes in anthocyanin coloration (Meyer & Saedler, 1996). In some cases even weak promoters have been used to silence genes. More recently PTGS has been suggested to be probably due to the involvement of a form of aberrant or double stranded RNA. Constructs containing repeats that specifically confer aberrant RNAs have been shown to be particularly useful for silencing (Hamilton et al., 1998; Waterhouse et al., 1998). By the transformation of specific gene constructs that confer aberrant or double stranded RNA quite predictable silencing of homologous genes in the genome is likely.

This approach of HDGS is also applicable to create mutations for multigene families, where clues to gene function may be obtained, although the interpretation of a phenotype might be complicated by suppression of the other homologous members. Suppression of gene activity in metabolic processes has been demonstrated in a number of cases and forms the basis for metabolic engineering that is applicable in transgenic crops (Herbers & Sonnewald, 1996; Ohlrogge, 1999). As a general HTP strategy for gene-function discovery it is too time consuming as several independent transformants need to be generated for every gene and have to be tested individually. Suppression of essential genes that give a dominant lethal phenotype also cannot be easily recovered and will escape attention.

A technology for silencing that lends itself to HTP methods is virus-induced gene silencing (VIGS). This employs virus vectors carrying fragments from the exons of plant host genes, that can be introduced by transformation or viral infection into the host and can result in the suppression of endogenous gene expression (Baulcombe, 1999). These suppressed phenotypes are phenocopies known mutations in host genes. The mechanism is based on an RNA-mediated defense against viruses and is also probably involved in PTGS achieved by transformation methods. At present VIGS involves the cloning of gene fragments of interest in vectors developed from the genomes of TMV, PVX or TGMV for applications in *Nicotiana* species. These systems can be stepped up for a HTP based forward genetics screen using libraries of gene fragments (e.g. cDNA) that are used for plant infection with pools or individual clones. Even members of multigene families that have sufficient DNA homology in fragments used will be effective in gene suppression. The recognition that VIGS and PTGS are related to antiviral

defense in plants and that viruses produce suppressors of this antiviral defense, offer clues to search for virus vectors suitable for VIGS in *Arabidopsis* or larger monocot genomes. The ability to use VIGS to obtain loss-of-function phenotypes, including for multigene families, will help close the phenotype gap.

Modified genes

Mutant proteins that disrupt the activity of the wild-type gene can confer dominant negative mutant phenotypes (Herskowitz, 1987). These may be by mutations in a regulatory protein that change it from an activator to a repressor, or mutations in the DNA binding domain of multimeric proteins so that non-functional mixed aggregates would compete the wild-type function. A number of dominant regulatory mutants have been isolated in plants like *C1-I* in maize inhibiting anthocyanin (Paz-Ares et al., 1990) and *ETR1* (Chang et al., 1993) in *Arabidopsis* showing insensitivity to ethylene.

A transgenic approach to manipulate and introduce dominant mutations was demonstrated for the *Arabidopsis* MADS box gene *AG*, where specific protein domains were overexpressed (Mizukami et al., 1996) by the 35S promoter displaying novel phenotypes, and providing an insight into protein regulatory mechanisms in flower development. Such mutant regulatory forms have also been shown to work when introduced into heterologous species (Mandel et al., 1992). By expression of an activated form of the LFY protein (LFY:VP16) the downstream controlled floral homeotic genes could be elucidated (Busch et al., 1999). The induction of transcription factors (Bruce et al., 2000) by estradiol-inducible systems helped uncover coordinately regulated genes and regulatory pathways. These examples reveal how gene/protein engineering can be useful to uncover functions. With our developing understanding of protein function, modification and interaction, methods of protein engineering will be increasingly used to make and test hybrid, shuffled and new proteins.

An integrated view of functional genomics

Classical genetics has tended to define mutants with distinct differences that are limited to a description of a handful of genes. Mutants that displayed weak and inconsistent effects were only poorly incorporated in the genetic models of pathways. In some genetic models,

256

quantitative trait loci have been recognized but have not yet been described by gene actions. Now, reverse genetics strategies, with the help of semi-quantitative measures of gene action, can offer new approaches to judge the effect of mutations. Mutants of redundant genes can also be combined to obtain multiple mutant phenotypes that reveal novel gene functions.

Mutants or expression variants of genes obtained by any of the various techniques described here, can be tested for changes in global gene expression monitored by the HTP methods for transcriptome, proteome or metabolome analysis. Regulatory-gene expression studies can help discover coordinately regulated pathways (Bruce et al., 2000) and using 'guilt-by-association' principles can uncover relationships between pathways. All these studies will contribute to a global genomic view of biology with interconnected genetic circuits, protein interactions and metabolic fluxes.

The phenotype of plants has been emphasized here primarily because this attribute is tangible and also applicable to trait development. Phenotype is beyond just the morphology of a mutant, but can also be described by gene expression manifested at the protein or metabolite level. The phenotype in its broadest context includes the effect of the expression of a gene, monitored by knockout or mis-expression, and its effect on the expression of other transcripts, proteins and metabolites.

Functional genomics is thus expanding the scope of biological investigation from a one-gene approach to a more systems-based holistic approach. The Phenome is the physical totality of all traits of an organism or of one of its subsystems (Mahner & Kary, 1997). This can now be described in our emerging integrated view of the genome, transcriptome, proteome and metabolome of an organism.

Genomic perspective

Worldwide about 40 million hectares of transgenic plants were estimated to be grown last year (Ohlrogge, 1999) and this might triple in the next five years. Over 90% of the commercial transgenic crops were engineered for either herbicide or pest tolerance and represent the first wave of crop engineering. These engineered traits provide lower production costs and/or higher yields. The emerging second wave of crop engineering will provide new or improved higher-value products for new markets and will include plant genes

with specific functions of commercial use. These genes are expected to be identified by genomic-based strategies.

In agricultural production a series of traits of importance in valuable crops like cereals, sugarcane, soybeans, cotton, potato, woody trees, horticultural and ornamental crops etc, can be addressed in a generic perspective using gene function discovery from model plants. The gene pool of crop plants has diverged over 150 million years and the diversity has arisen from variations in the basic building blocks of genes determining specific plant processes. The isolation of genes determining plant height from *Arabidopsis* led to the identification of orthologous (similar and same function) dwarf genes in rice and other cereals (Peng et al., 1999). Most surprisingly these are the dwarf genes that steered in the Green Revolution, that have been identified and incorporated by classic breeding programs.

Many genes for important traits in metabolic engineering are conserved and transferable between model plants and crops. Some regulatory gene functions are also conserved, as demonstrated with *LEAFY* from *Arabidopsis* that can function in plants as diverse as aspen trees and rice for manipulation of early flower initiation (Weigel & Nilsson, 1995). Although the fine-tuning of such trait modification remains to be done, the first steps come from an understanding of the basic biology in the model plant. The next step is to analyze such a trait on a whole genome analysis, by identifying the interaction with other genes in the pathway. The identification of a few key genes from *Arabidopsis* for generic traits like resistance to the abiotic stresses cold, salt and drought can also lead to their application in diverse crop plants that are difficult to study because of their complex genomes and growth habits.

The generation of HTP genome and EST sequence information requires large capital investments that have been made by a number of private sector enterprises with expectations of large returns in the future. In the light of the genomics revolution the emphasis is changing from Chemicals, Pharmaceuticals and Agribusiness towards biotechnology and the life sciences industry. With all the top life science players involved in mergers, acquisitions and alliances all vying for life science products, including plant biotechnology, a lot of intellectual property generated by patenting genomic-based information is gathering momentum. The race to own information, protect and use it is on. This large-scale creation and protection of knowledge

ese
sed

im-
oy-
ınd
:ric
del
ver
om
de-
of
led
me
:ng
'arf
ave
ing

en-
:en
nc-
FY
di-
ırly
ıgh
· be
· of
ɔ is
sis,
the
om
ıbi-
to
:ult
vth

ıce
hat
:er-
ıre.
s is
Ag-
ces
/ed
for
·gy,
ing
ım.
on.
lge

is considered to be the central factor for improving world health and living standards.

Genomics can offer us the new Green Revolution to solve the world's hunger problem. To achieve this the race in patenting and using genome information will have to give way to a sustainable strategic utilization in the developing world. Meanwhile, systematic plant genome function analysis will provide us with information that will revolutionize our view of looking at plants and crop production. Most importantly, our understanding of the genome will make the manipulation of transgenic traits more predictable and provide grounds for consumer confidence.

Acknowledgements

References

Aarts MGM, Dirkse W, Stiekema WJ and Pereira A (1993) Transposon tagging of a male sterility gene in Arabidopsis. Nature 363: 715–717.

Arondel V, Lemieux B, Hwang I, Gibson S, Goodman HM and Somerville CR (1992) Map-based cloning of a gene controlling omega-3 fatty acid desaturation in Arabidopsis. Science 258: 1353–1355.

Azpiroz-Leehan R and Feldmann KA (1997). T-DNA insertion mutagenesis in Arabidopsis: going back and forth. Trends Genetics 13: 152–156.

Babiychuk E, Fuangthong M, van Montague M, Inze D and Kushnir S (1997) Efficient gene tagging in Arabidopsis thaliana using a gene trap approach. Proc Natl Acad Sci USA 94: 12722–12727.

Baker B, Schell J, Lörz H and Fedoroff N (1986) Transposition of the maize controlling element 'Activator' in tobacco. Proc Natl Acad Sci USA 83: 4844–4848.

Baker B, Coupland G, Fedoroff N, Starlinger P and Schell J (1987) Phenotypic assay for excision of the maize controlling element Ac in tobacco. EMBO J 6: 1547–1554.

Baldwin D, Crane V and Rice D (1999) A comparison of gel-based, nylon filter and microarray techniques to detect differential RNA expression in plants. Curr Opin Plant Biol 2: 96–103.

Ballinger DG and Benzer S (1989). Targeted gene mutations in Drosophila. Proc Natl Acad Sci USA 86: 9402–9406.

Bancroft I and Dean C (1993). Transposition pattern of the maize element Ds in Arabidopsis thaliana. Genetics 134: 1221–1229.

Bancroft I, Bhatt AM, Sjodin C, Scofield S, Jones JDG and Dean C (1992) Development of an efficient two-element transposon tagging system in Arabidopsis thaliana. Mol Gen Genet 233: 449–461.

Bancroft I, Jones JDG and Dean C (1993). Heterologous transposon tagging of the DRL1 locus in Arabidopsis. Plant Cell 5: 631–638.

Baulcombe DC (1999) Fast forward genetics based on virus-induced gene silencing. Curr Opin Plant Biol 2: 109–113.

Bechtold N and Pelletier G (1998) In planta Agrobacterium-mediated transformation of adult Arabidopsis thaliana plants by vacuum infiltration. Methods Mol Biol 82: 259–266.

Bouchez D and Höfte H (1998) Functional genomics in plants. Plant Physiol 118: 725–732.

Brenner S, Williams SR, Vermaas EH, Storck T, Moon K, McCollum C, et al. (2000) In vitro cloning of complex mixtures of DNA on microbeads: Physical separation of differentially expressed cDNAs. Proc Natl Acad Sci USA 97: 1665–1670.

Bruce W, Folkerts O, Garnaat C, Crasta O, Roth B and Bowen B (2000) Expression profiling of the maize flavanoid pathway genes controlled by estradiol-inducible transcription factors CRC and P. Plant Cell 12: 65–80.

Burns N, Grimwade R, Macdonald PB, Choi E-Y, Finberg K, Roeder GS et al. (1994) Large-scale analysis of gene expression, protein localization, and gene disruption in Saccharomyces cerevisiae. Genes Dev 8: 1087–1105.

Busch MA, Bomblies K and Weigel D (1999) Activation of a floral homeotic gene in Arabidopsis. Science 285: 585–587.

Campisi L, Yang Y, Yi Y, Heilig E, Herman B, Cassista AJ, et al. (1999) Generation of enhancer trap lines in Arabidopsis and characterization of expression patterns in the inflorescence. Plant J 17: 699–707.

Chang C, Kwok SF, Bleecker AB and Meyerowitz EM (1993) Arabidopsis ethylene-response gene ETR1: similarity of product to two-component regulators. Science 262: 539–544.

Chin HG, Choe MS, Lee S-H, Park SH, Park SH, Koo JC, et al. (1999) Molecular analysis of rice plants harboring an Ac/Ds transposable element-mediated gene trapping system. Plant J 19: 615–623.

Cho RJ, Mindrinos M, Richards DR, Sapolsky RJ, Anderson M, Drenkard E, et al. (1999) Genome-wide mapping with biallelic markers in Arabidopsis thaliana. Nat Genet 23: 203–207.

Clough SJ and Bent AF (1998) Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. Plant J 16: 735–743.

Coen ES, Robbins TP, Almeida J, Hudson A and Carpenter R (1989) Consequences and mechanisms of transposition in Antirrhinum majus. In: Berg DE and Howe MM (eds.) Mobile DNA, (pp. 413–436) Am. Soc. Microbiol, Washington DC.

Cooke R, Raynal M, Laudié M and Delseny M (1997) Identification of members of gene families in Arabidopsis thaliana by contig construction from partial cDNA sequences: 106 genes encoding 50 cytoplasmic ribosomal proteins. Plant J 11: 1127–1140.

Das L and Martienssen R (1995) Site-selected transposon mutagenesis at the hcf106 locus in maize. Plant Cell 7: 287–294.

Desprez T, Amselem J, Caboche M and Höfte H (1998) Differential gene expression in Arabidopsis monitored using cDNA arrays. Plant J 14: 643–652.

Dove A (1999) Proteomics: translating genomics into products. Nat Biotech 17: 233–236.

Dubois P, Cutler S and Belzile FJ (1998) Regional insertional mutagenesis on chromosome III of Arabidopsis thaliana using the maize Ac element. Plant J 13: 141–151.

Enoki H, Izawa T, Kawahara M, Komatsu M, Koh S, Kyozuka J et al. (1999) Ac as a tool for the functional genomics of rice. Plant J 19: 605–613.

Fedoroff NV and Smith DL (1993) A versatile system for detecting transposition in Arabidopsis. Plant J 3: 273–289.

258

Feldmann KA (1991) T-DNA insertion mutagenesis in *Arabidopsis*: mutational spectrum. *Plant J* 1: 71–82.

Gale MD and Devos KM (1998) Plant comparative genetics after 10 years. *Science* 282: 656–659.

Gaymard F, Pilot G, Lacombe B, Bouchez D, Bruneau D, Boucherez J, et al. (1998) Identification and disruption of a plant Shaker-like outward channel involved in K+ release into the xylem sap. *Cell* 94: 647–655.

Gerats AGM, Huits H, Vrijlandt E, Maraña C, Souer E and Beld M (1990) Molecular characterization of a non-autonomous transposable element (*dTph1*) of petunia. *Plant Cell* 2: 1121–1128.

Hamilton AJ, Brown S, Yuanhai H, Ishizuka M, Lowe A, Solis A-GA et al. (1998) A transgene with repeated DNA causes high frequency, post-transcriptional suppression of ACC-oxidase gene expression in tomato. *Plant J* 15: 737–746.

Herbers K and Sonnewald U (1996) Manipulating metabolic partitioning in transgenic plants. *Trends Biotech* 14: 198–205.

Herskowitz I (1987) Functional inactivation of genes by dominant negative mutations. *Nature* 329: 219–222.

Hirochika H (1997) Retrotransposons of rice: their regulation and use for genome analysis. *Plant Mol Biol* 35: 231–240.

Hirsch RE, Lewis BD, Spalding EP and Sussman MR (1998) A role for AKT1 potassium channel in plant nutrition. *Science* 280: 918–921.

Höfte H, Desperez T, Amselem J, Chiapollo H, Caboche M et al. (1993) An inventory of 1152 expressed sequence tags obtained by partial sequencing of cDNAs from *Arabidopsis thaliana*. *Plant J* 4: 1051–1061.

Ito T, Seki M, Hayashida N, Shibata D and Shinozaki K (1999) Regional insertional mutagenesis of genes on *Arabidopsis thaliana* chromosome V using the *Ac/Ds* transposon in combination with a cDNA scanning method. *Plant J* 17: 433–444.

Jones DA, Thomas CM, Hammond-Kosack KE, Balint-Kurti PJ and Jones JDG (1994) Isolation of the tomato *Cf-9* gene for resistance to *Cladosporium fulvum* by transposon tagging. *Science* 266: 789–793.

Jones JDG, Carland FM, Maliga P and Dooner HK (1989) Visual detection of transposition of the maize element *Activator* (*Ac*) in tobacco seedlings. *Science* 244: 204–207.

Jones JDG, Carland FM, Lin E, Ralston E and Dooner HK (1990) Preferential transposition of the maize element *Activator* to linked chromosomal locations in tobacco. *Plant Cell* 2: 701–707.

Kaiser K and Goodwin SF (1990) Site-selected transposon mutagenesis of *Drosophila*. *Proc Natl Acad Sci USA* 87: 1686–1690.

Kakimoto T (1996) CKI1, a histidine kinase homolog implicated in cytokinin signal transduction. *Science* 274: 982–985.

Kardailsky I, Shukla VK, Ahn JH, Dagenais N, Christensen SK, Nguyen JT, et al. (1999) Activation tagging of the floral inducer *FT*. *Science* 286: 1962–1965.

Kertbundit S, De Greve H, De Boeck F, van Montague M and Hernalsteens JP (1991) *In vivo* random ß-glucuronidase gene fusions in *Arabidopsis thaliana*. *Proc Natl Acad Sci USA* 88: 5212–5216.

Kobayashi Y, Kaya H, Goto K, Iwabuchi M and Araki T (1999) A pair of related genes with antagonsitic roles in mediating flowering signals. *Science* 286: 1960–1962.

Koes R, Souer E, van Houwelingen A, Mur L, Spelt C, Quattrocchio F, et al. (1995) Targeted gene inactivation in petunia by PCR-based selection of transposon insertion mutants. *Proc Natl Acad Sci USA* 92: 8149–8153.

Koncz C, Martini N, Mayerhofer R, Koncz-Kalman Zs, Körber H, Redei GP et al. (1989) High-frequency T-DNA-mediated gene tagging in plants. *Proc Natl Acad Sci USA* 86: 8467–8471.

Koncz C, Nemeth K, Redei GP and Schell J (1992) T-DNA insertional mutagenesis in *Arabidopsis*. *Plant Mol Biol* 20: 963–976.

Kooter JM, Matzke MA and Meyer P (1999) Listening to the silent genes: transgene silencing, gene regulation and pathogen control. *Trends Plant Sci* 4: 340–347.

Krysan PJ, Young JF, Tax F and Sussman MR (1996) Identification of transferred DNA insertions within *Arabidopsis* genes involved in signal transduction and ion transport. *Proc Natl Acad Sci USA* 93: 8145–8150.

Krysan PJ, Young JC and Sussman MR (1999) T-DNA as an insertional mutagen in *Arabidopsis*. *Plant Cell* 11: 2283–2290.

Lemieux B, Aharoni A and Schena M (1998) Overview of DNA chip technology. *Mol Breeding* 4: 277–289.

Lin X, Kaul S, Rounsley S, Shea TP, Maria-Ines B et al. (1999) Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*. *Nature* 402: 761–768.

Liu LX, Spoerke JM, Mulligan EL, Chen J, Reardon B et al. (1999) High-throughput isolation of *Caenorhabditis elegans* deletion mutants. *Genome Res* 9: 859–867.

Long D, Martin M, Sundberg E, Swinburne J, Puangsomlee P and Coupland G (1993) The maize transposable element system *Ac/Ds* as a mutagen in *Arabidopsis*: Identification of an *albino* mutation induced by *Ds* insertion. *Proc Natl Acad Sci USA* 90: 10370–10374.

Mahner M and Kary M (1997) What exactly are genomes, genotypes and phenotypes? And what about phenomes? *J Theor Biol* 186: 55–63.

Mandel MA, Bowman JL, Kempin SA, Ma H, Meyerowitz EM and Yanofsky MF (1992) Manipulation of flower structures in transgenic tobacco. *Cell* 71: 133–143.

Martienssen RA (1998) Functional genomics: Probing plant gene function and expression with transposons. *Proc Natl Acad Sci USA* 95: 2021–2026.

Masson P and Fedoroff N (1989) Mobility of the maize *Suppressor-mutator* element in transgenic tobacco cells. *Proc Natl Acad Sci USA* 86: 2219–2223.

Mathur J, Szabados L, Schaefer S, Grunenberg B, Lossow A, Jonas-Straube E, et al. (1998) Gene identification with sequenced T-DNA tags generated by transformation of *Arabidopsis* cell suspension. *Plant J* 13: 707–716.

Matsumura H, Nirasawa S and Terauchi R (1999) Transcript profiling in rice (*Oryza sativa* L.) seedlings using serial analysis of gene expression (SAGE). *Plant J* 20: 719–726.

Mayer K, Schuller C, Wambutt R, Murphy G, Volckaert G et al. (1999) Sequence and analysis of chromosome 4 of the plant *Arabidopsis thaliana*. *Nature* 402: 769–777.

McKinney EC, Ali N, Traut A, Feldmann KA, Belostotsky DA, McDowell JM et al. (1995) Sequence-based identification of T-DNA insertion mutations in *Arabidopsis* actin mutants *act2-1* and *act4-1*. *Plant J* 8: 613–622.

Meissner RC, Jin H, Cominelli E, Denekamp M, Fuertes A, Greco R, et al. (1999) Function search in a large transcription factor gene family in *Arabidopsis*: Assessing the potential of reverse genetics to identify insertional mutations in R2R3 *MYB* genes. *Plant Cell* 11: 1827–1840.

Mewes HW, Albermann K, Bähr M, Frishman D, Gleissner A, Hani J, et al. (1997) Overview of the yeast genome. *Nature* 387: 7–9.

Meyer P and Saedler H (1996) Homology dependent gene silencing in plants. *Ann Rev Plant Physiol Plant Mol Biol* 47: 23–48.

Meyers BC, Dickerman AW, Michelmore RW, Sivaramakrishnan S, Sobral BW and Young ND (1999) Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding superfamily. *Plant J* 20: 317–332.

inser-
-976.
silent
ontrol.

cation
volved
i USA

as an
290.
DNA

1999)
dopsis

1999)
:letion

ilee P
ystem
albino
iA 90:

geno-
r Biol

z, EM
res in

t gene
ad Sci

essor-
ad Sci

Jonas-
enced
s cell

it pro-
sis of

et al.
plant

DA,
of T-
ict2-1

Greco
factor
everse
genes.

er A,
387:

ncing

ian S,
genes
within

Miklos GLG and Rubin GM (1996) The role of the genome project in determining gene function: insights from model organisms. Cell 86: 521–529.

Mizukami Y, Huang H, Tudor M, Hu Y and Ma H (1995) Functional domains of the floral regulator AGAMOUS: Characterization of the DNA binding domain and analysis of dominant negative mutations. Plant Cell 8: 831–845.

Mol J, Grotewold E and Koes R (1998) How genes paint flowers and seeds. Trends Plant Sci 3: 212–217.

Nam YW, Penmetsa RV, Endre G, Uribe P, Kim D and Cook DR (1999) Construction of a bacterial artificial chromosome library of Medicago truncatula and identification of clones carrying ethylene-response genes. Theor Appl Genet 98: 638–646.

Neff MM, Nguyen SM, Malancharuvil EJ, Fujioka S, Noguchi T, Seto H, et al. (1999) BAS1: A gene regulating brassinosteroid levels and light responsiveness in Arabidopsis. Proc Natl Acad Sci USA 96: 15316–15323.

Newman T, de Bruijn FJ, Green P, Keegstra K, Kende H, McIntosh L, et al. (1994) Genes galore: a summary of methods for accessing results from large-scale partial sequencing of anonymous Arabidopsis cDNA clones. Plant Physiol 106: 1241–1255.

Ohlrogge J (1999) Plant metabolic engineering: are we ready for phase two? Curr Opin Plant Biol 2: 121–122.

O'Keefe DP, Tepperman JM, Dean C, Leto KJ, Erbes DL and Odell JT (1994) Plant expression of a bacterial cytochrome P450 that catalyzes activation of a sulfonylurea pro-herbicide. Plant Physiol 105: 473–482.

Osborne BI, Wirtz U and Baker B (1995) A system for insertional mutagenesis and chromosomal rearrangement using the Ds transposon and Cre-lox. Plant J 7: 687–701.

Parinov S, Sevugan M, Ye D, Yang W-C, Kumaran M and Sundaresan V (1999) Analysis of flanking sequences from dissociation insertion lines: A database for reverse genetics in Arabidopsis. Plant Cell 11: 2263–2270.

Paz-Ares J, Ghosal D and Saedler H (1990) Molecular analysis of the C1-I allele from Zea mays: A dominant mutant of the regulatory C1 locus. EMBO J 9: 315–321.

Peng J, Richards DE, Hartley NM, Murphy GP, Devos KM, Flintham JE, et al. (1999) 'Green revolution' genes encode mutant gibberellin response modulators. Nature 400: 256–261.

Pereira A (1998) Heterologous transposon tagging systems. In: Lindsey K (ed) Transgenic Plant Research (pp. 91–108) Harwood Academic Publishers, UK.

Pereira A and Saedler H (1989) Transpositional behaviour of the maize En/Spm element in transgenic tobacco. EMBO J 8: 1315–1321.

Quackenbush J, Liang F, Holt I, Pertea G and Upton J (2000) The TIGR gene indices: reconstruction and representation of expressed gene sequences. Nuc Acids Res 28: 141–145.

Rabinowicz PD, Schutz K, Dedhia N, Yordan C, Parnell LD, Stein L, et al. (1999) Differential methylation of genes and retrotransposons facilitates shotgun sequencing of the maize genome. Nat Genet 23: 305–308

Rounsley SD, Glodek A, Sutton G, Adams MD, Somerville CR, Venter JC et al. (1996) The construction of Arabidopsis expressed sequence tag assemblies. Plant Physiol 112: 1177–1183.

Ruan Y, Gilmore J and Conner T (1998) Towards Arabidopsis genome analysis: monitoring expression profiles of 1400 genes using cDNA microarrays. Plant J 15: 821–823.

Schena M, Shalon D, Davis RW and Brown PO (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 270: 467–470.

Schmidt R, West J, Love K, Lenehan Z, Lister C, Thompson H et al. (1995) Physical map and organization of Arabidopsis thaliana chromosome 4. Science 270: 480–483

Shirasu K, Lahaye T, Tan M-W, Zhou F, Azevedo C and Schulze-Lefert P (1999) A Novel class of eukaryotic zinc-binding proteins is required for disease resistance signaling in barley and development in C. elegans. Cell 99: 355–366.

Skarnes WC (1990) Entrapment vectors: a new tool for mammalian genetics. Biotechnology 8: 827–831.

Smith DL and Fedoroff NV (1995) LRP1, a gene expressed in lateral and adventitous root primordia of Arabidopsis. Plant Cell 7: 735–745.

Solano R and Ecker JR (1998) Ethylene gas: perception, signaling and response. Curr Opin Plant Biol 1: 393–398.

Speulman E, Metz PLJ, van Arkel G, te Lintel Hekkert B, Stiekema WJ and Pereira A (1999) A two-component enhancer-inhibitor transposon mutagenesis system for functional analysis of the Arabidopsis genome. Plant Cell 11: 1853–1866.

Springer PS, McCombie WR, Sundaresan V and Martienssen RA (1995) Gene trap tagging of PROLIFERA, an essential MCM2-3-5-like gene in Arabidopsis. Science 268: 877–880.

Stuurman J, Nijkamp HJJ and van Haaren MJJ (1998) Molecular insertion-site selectivity of Ds in tomato. Plant J 14: 215–223.

Sundaresan V, Springer P, Volpe T, Haward S, Jones JDG, Dean C, et al. (1995) Patterns of gene action in plant development revealed by enhancer trap and gene trap transposable elements. Genes Dev 9: 1797–1810.

Swinburne J, Balcells L, Scofield SR, Jones JDG and Coupland G (1992) Elevated levels of Activator transposase mRNA are associated with high frequencies of Dissociation excision in Arabidopsis. Plant Cell 4: 583–595.

The C. elegans sequencing consortium (1998) Genome sequence of the nematode C.elegans: a platform for investigating biology. Science 282: 2012–2018.

Tissier AF, Marillonnet S, Klimyuk V, Patel K, Torres MA, Murphy G et al. (1999) Multiple independent defective suppressor-mutator transposon insertions in Arabidopsis: A tool for functional genomics. Plant Cell 11: 1841–1852.

Topping JF and Lindsey K (1995) Insertional mutagenesis and promoter trapping in plants for the isolation of genes and the study of development. Transgenic Res 4: 291–305.

Trethewey RN, Krotzky AJ and Willmitzer L (1999) Metabolic profiling: a Rosetta Stone for genomics. Curr Opin Plant Biol 2: 83–85.

Tsugeki R, Kochieva EZ and Fedoroff NV (1996) A transposon insertion in the Arabidopsis SSR16 gene causes an embryo-defective lethal mutation. Plant J 10: 479–489.

van Haaren MJJ and Ow DW (1993) Prospects of applying a combination of DNA transposition and site-specific recombination in plants: a strategy for gene identification and cloning. Plant Mol Biol 23: 525–533.

Walbot V (1992) Strategies for mutagenesis and gene cloning using transposon tagging and T-DNA insertional mutagenesis. Annu Rev Plant Physiol & Plant Mol Biol 43: 49–82.

Waterhouse PM, Graham MW and Wang M-B (1998) Virus resistance and gene silencing in plants can be induced by simultaneous expression of sense and antisense RNA. Proc Natl Acad Sci USA 95: 13959–13964.

Weigel D and Nilsson O (1995) A developmental switch sufficient for flower initiation in diverse plants. Nature 377: 495–500.

Weigel D, Ahn JH, Blazquez MA, Borevitz JO, Christensen SK, Fankhauser C et al. (2000) Activation tagging in Arabidopsis Plant Physiol 122: 1003–1013.

260

Wilson K, Long D, Swinburne J and Coupland G (1996) A *dissociation* insertion causes a semidominant mutation that increases expression of *TINY*, an *Arabidopsis* gene related to *APETALA2*. *Plant Cell* 8: 659–671.

Winkler RG, Frank MR, Galbraith DW, Feyereisen R and Feldmann KA (1998) Systematic reverse genetics of transfer-DNA-tagged lines of *Arabidopsis*. *Plant Physiol* 118: 743–750.

Wisman E, Hartmann U, Sagasser M, Baumann E, Palme K, Hahlbrock K, et al. (1998) Knock-out mutants from an *En-1* mutagenized *Arabidopsis thaliana* population generate phenylpropanoid biosynthesis phenotypes. *Proc Natl Acad Sci USA* 95: 12432–12437.

Yoder JI (1990) A genetic analysis of mutations recovered from tomato following *Agrobacterium*-mediated transformation with the maize transposable elements *Activator* and *Dissociation*. *Theor Appl Genet* 79: 657–662.

Zwaal RR, Broeks A, van Meurs J, Groene JTM and Plasterk RHA (1993) Target-selected gene inactivation in *Caenorhabditis elegans* by using a frozen transposon insertion mutant bank. *Proc Natl Acad Sci USA* 90: 7431–7435.